

## An Investigation of Predictors of Information Diffusion in Social Media: Evidence from Sentiment Mining of Twitter Messages

Mohammad Salehan  
Cal Poly Pomona  
[msalehan@cpp.edu](mailto:msalehan@cpp.edu)

Dan J. Kim  
University of North Texas  
[dan.kim@unt.edu](mailto:dan.kim@unt.edu)

### Abstract

*Social media have facilitated information sharing in social networks. Previous research shows that sentiment of text influences its diffusion in social media. Each emotion can be located on a three-dimensional space formed by dimensions of valence (positive-negative), arousal (passive / calm-active / excited), and tension (tense-relaxed). While previous research has investigated the effect of emotional valence on information diffusion in social media, the effect of emotional arousal remains unexplored. This study examines how emotional arousal influences information diffusion in social media using a sentiment mining approach. We propose a research model and test it using data collected from Twitter.*

### 1. Introduction

Social media have significantly changed the way humans communicate. Many people use social media to keep in touch with family and friends and receive up-to-date information about what happens around the world. In recent years, social media have experienced tremendous growth in the number of users. Facebook alone has more than 1.3 billion active users [1] and Twitter has attracted more than 600 million active users [2]. Social media have been used to support political campaigns of the candidates of US presidential elections [3]. Politicians are now using Twitter as tool for public diplomacy and to release the up-to-date progress of their negotiations [4].

Social media have facilitated information sharing in social networks. Previous research shows that several factors including content related factors and user and network characteristics influence information diffusion in social media. Content related factors include topic [5], URL, hashtags [6], and user and network characteristics including social capital perception [7], popularity, and homophile [8]. While previous research has shown that computer-mediated communications (CMC) can effectively transfer emotions, recent research has focused on the effect of emotions on information diffusion in social media.

The emotions contained in a computer-mediated message significantly influence how the message is processed and interpreted by the receiver [9-12]. Similarly, sentiment of a messages may influence its diffusion in social media [13].

Each emotion can be located on a three-dimensional space formed by dimensions of *valence* (positive – negative), *arousal* (passive / calm – active / excited), and *tension* (tense-relaxed) [14]. These three dimensions covary with physical states of the body such as physiological arousal. While previous research has noted the effect of emotions on information dissemination on social media [13], many aspects of the problem remain unexplored. Among the three dimensions of emotions, valence has received the highest level of attention. Most studies that utilize sentiment mining to study social media, solely focus on emotional valence and total amount of sentiment in the text. For example, higher levels of total sentiment in a tweet (i.e., both positive and negative) are related to its retweet performance [13].

However, very few studies investigate the effect of emotional arousal on information diffusion. Emotional arousal influences individuals in several ways. Emotional arousal has been shown to increase action-related behaviors such as moving to help others [15], to influence the decision making process of individuals [16], and to affect lexical decision response times [17, 18]. Considering the importance of emotional arousal on human behavior, this study considers emotional arousal as a predictor of information diffusion in social media.

This study aims to extend our knowledge on information diffusion in social media by analyzing the diffusion performance of messages in Twitter which is a powerful tool for information sharing. Information diffusion in Twitter is heavily dependent on retweets. Retweet activity reflects how the social network directs the propagation of information [19]. The more a message is retweeted, the more people will view it and it becomes more likely to get into trending topics. Thus, we use retweet count as the performance measure of information diffusion in Twitter. Using sentiment mining as an approach for data analysis, we

suggest that emotions significantly predict information diffusion in social media. We propose a research model that explains the relationship between different types of sentiment and information diffusion. Then, we test the proposed research model using the data collected from Twitter.com website and discuss the findings and implications. The findings of the study will help individuals, news broadcasting agencies, politicians, and mass media campaigns to improve their performance on social media.

## **2. Literature review and theoretical background**

### **2.1. Information diffusion in social media**

Diffusion of information in social media has been studied in several disciplines including physical, social, and computational sciences. In Business and Marketing research, information diffusion has also been studied with different titles such as electronic word of mouth and viral marketing [e.g., 11, 20]. While different types of social media platforms such as social networking services [e.g., 21], online review websites [e.g., 12], photo sharing websites [e.g., 22], weblogs [e.g., 23], and online communities [e.g., 24] are used by researchers to study information diffusion, Twitter has received significant attention from academics because of the way it facilitates diffusion of information in the form of retweets [e.g., 25, 26-31]. Twitter is particularly important in this context because 75% of Twitter users mainly use it to access information [7].

Previous research looks at both quantity as well as speed of retweeting as performance measures for information diffusion in Twitter [13]. By analyzing over a million tweets, Nagarajan, Purohit [5] found that users are more likely to retweet than get involved in direct conversation. They showed that popular tweets fall in on the four categories of call for social action, collective group identity-making, crowdsourcing, and information sharing; the tweets from the first three categories create a sparse network (i.e., loosely connected) while the last category, information sharing, has a dense retweet network.

By Analyzing 10000 tweets, Suh, Hong [6] examined how content features (such as URL inclusion, hashtags, and mentions) and contextual features (including the number of followers and followees, the age of the account, the number of favorited tweets, and the number and frequency of tweets influence diffusion of information in Twitter. They found that tweets containing URL and/or hashtag are more likely to be retweeted. Number of

followers/followees and age of the account are also positively related to retweet count. The tweets mentioning other user(s) are less likely to be retweeted.

Macskassy and Michelson [8] used snowball sampling to collect around 11432 Twitter user accounts from which they collected over 353000 tweets. Thirty-two percent of the messages were retweets. They analyzed and compared four different models to predict retweet behavior including random model (as benchmark), recent communication model (retweet those recently been in contact with), topic model (retweet topic of interest), and homophily (profile) model (retweet those with similar profiles). They found that the homophily model outperformed the other models followed by recent mode, topic model, and random model.

Recuero, Araujo [7] investigated the effect of social capital on retweet behavior. They found that retweeting not only benefits social network as a whole by spreading information, it but also benefits individual users by allowing them to reach those out of their network while their identity is attached to the message. Because information access is an important motivation for retweeting, timing is an important aspect of retweets. They also found that mentioning the original source when retweeting can add credibility to the message. However, they showed that messages that contain too many mentions may look old and hence many users cut the number of mentions so that the information looks fresh; many users retweet because it is a convenient way of feeding their followers without actually producing any information. They also found that some users retweet certain people in order to demonstrate their social network. Retweet behavior is also a form of agreement with an expressed idea.

### **2.2. Emotions in CMC**

Previous research has indicated that CMC can effectively transfer emotions. The receiver of a message can detect the sender's emotions through verbal cues such as emotion words as well as nonverbal cues such as emoticons [32]. Moreover, the emotions contained in a message transferred through CMC significantly influence how the message is processed and interpreted by the receiver [9, 10]. Sentiment mining can be used to extract sentiment from the text and consequently to predict the behavior of the receiver of the message [12, 33-35].

Different categorizations of emotions exist in the literature. Wundt [14] proposed a dimensional approach for classifying emotions. He suggested each emotion can be located on a three-dimensional space

formed by dimensions of valence (positive–negative), arousal (calm–excited), and tension (tense–relaxed). He believed that these three dimensions covary with physical states of the body such as physiological arousal. Examples of emotions with high arousal and high valence include ambitious, adventurous, self-confident, and delighted. These emotions are all examples of positive emotions that are high in arousal. In the opposite corner is the low valence and low arousal section, containing bored, sad, depressed, and doubtful as some examples.

Mehrabian and Russell [36] proposed a three dimensional model of emotions similar to that of Wundt [14] composed of pleasure, arousal, and dominance. The model is called PAD emotional state model. The pleasure-displeasure scale represents how pleasant or unpleasant one feels about something. This dimension is similar to what Wundt [14] calls valence. The arousal dimension is the same as what Wundt [14] suggested. Finally, the dominance-submissiveness scale measures how controlling and dominant versus controlled or submissive one feels.

Plutchik [37] proposed a wheel of emotions consisting of 8 basic emotions and 8 advanced emotions each composed of 2 basic ones. The basic emotions include joy, trust, anticipation, fear, surprise, sadness, disgust, and anger. The advanced emotions include optimism, love, submission, awe, disapproval, remorse, contempt, and aggressiveness. As an example, anger is a highly unpleasant, very aroused, and moderately dominant emotion, while boredom is a little unpleasant, fairly unaroused, and mostly non-dominant.

### 2.3. Negativity bias

Prior research shows that people respond differently to positive and negative stimuli, and negative events tend to provoke stronger and faster emotional, behavioral, and cognitive reactions than neutral or positive events [38, 39]. More specifically, it has been shown that people are subject to a general bias to show greater weight to negative entities such as events, objects, and personal traits [38]. This is generally referred to as “negativity bias” [39]. Recent studies of communication in the social media context such as Facebook also show that negative sentiment postings induce more feedback in terms of comments compared to those with positive sentiment [25].

### 2.4. Emotional arousal

Different levels of emotional arousal have different impact on human behavior. Low levels of arousal are characterized by staying calm and relaxed.

Higher levels of arousal, in contrast, are characterized by activity [40]. High levels of arousal increase action-related behaviors such as moving to help others [15] and lead to faster response to offers in negotiations [41]. Stories with high levels of emotional arousal trigger greater emotional reaction in individuals and thus enhance their memory for the story [42].

Emotional arousal also influences lexical decision tasks. Lexical decision task is an experiment that involves measuring how quickly people identify stimuli as words or nonwords. A mixture of words and logatomes or pseudowords (nonsense strings that respect the phonotactic rules of a language, like *trud* in English) are presented to subjects. The goal is to recognize whether the presented stimulus is a word or not. Response times of subjects are measured as the dependent variable in the experiment.

Hofmann et al. [17] ran a series of experiments to investigate the effect of emotional arousal on lexical decision response times. They varied emotional valence and emotional arousal as experimental factors. They presented negative high-arousing and low-arousing words to subjects and measured their response times. They observed that high-arousing negative words are processed faster than both low-arousing negative words and neutral words. In contrast, low-arousal negative words are processed slower than neutral words. They also observed that negative valence does not influence response times when the treatment is controlled for emotional arousal. Finally, they found out that the effect of emotional arousal on response times is only valid for negative words but not for positive ones. Positive words improve the response time irrespective of their level of arousal. Another study finds that three discrete emotions of happiness, fear, and disgust significantly predict response times of lexical decision tasks. These three emotions describe as much variance in response times as dimensional and categorical models of emotions [43].

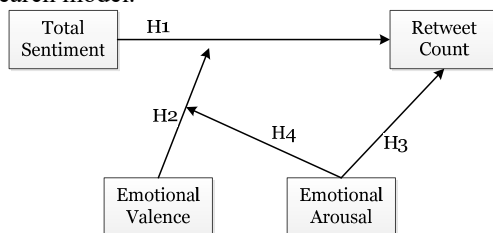
Several lexicons have been created for measurement of emotional arousal. A lexicon is a word list where each word has been scored for sentiment strength. The Affective Norms for English Words (ANEW), developed and distributed by the Center for Emotion and Attention (CSEA) at the University of Florida, is a lexicon based on PAD emotional state model [36] which has measured pleasure, arousal, and dominance for a set of English words [44]. The initial version included 150 words from Mehrabian and Russell [36] and 450 words from Bellezza, Greenwald [45] which together with other words the authors added created a list of 1040 words. The authors used undergrad psychology students to rate each word in terms of the three PAD dimensions

on a scale from 1 to 9 [44]. ANEW reports mean and standard deviation of each type of sentiment for each word. The latest version is more comprehensive and contains sentiment of 2476 words [46]. The lexicon has been extensively used and validated by previous research [47-52].

Whissell's Dictionary of Affect in Language is another lexicon developed by University of Columbia scholars [53]. The lexicon contains pleasantness (or valence), activation (or arousal), and imagery for 8742 English words. The authors used 200 volunteers, mostly university students, to rate each word on a three-point scale. The scale for pleasantness includes: (1) unpleasant, (2) in between, and (3) pleasant. Similarly, the scale for activation is: (1) passive, (2) in between, and (3) active. Finally, there are similar scale points for imagery including (1) hard to imagine, (2) in between, and (3) easy to imagine [54].

### 3. Research model and hypothesis

Building upon the previous literature on emotions, we propose a research model that describes the effect of sentiment on information diffusion on social media. We argue that the levels of total sentiment, emotional valence, and emotional arousal expressed in a Twitter message significantly influence its level of diffusion. Figure 1 shows our proposed research model.



**Figure 1. Research Model**

Sentiment of a message can be effectively communicated through the text and significantly influences the perceptions of the reader [9, 10, 32]. Moreover, previous research shows that affective language in the online environment receives more attention and feedback compared to neutral language. For example, people who use affective language in discussion forums receive more feedback than those who do not [55]. Both positive and negative sentiment increase participation but in different forms. Positive sentiment increases continued participation while negative sentiment triggers hostile interactions [56]. Hence, we expect the total amount sentiment (either positive or negative) expressed in a Twitter message to be positively related to the amount of feedback and attention it receives. A popular way of showing

attention toward a Twitter message is to retweet it [7]. Thus, we expect tweets with higher amount of sentiment to have a larger number of retweets. Consequently, we hypothesize that:

H1: The larger the total amount of sentiment (positive or negative) a Twitter message exhibits, the more often it is retweeted.

According to negativity bias, people react differentially in response to positive and negative stimuli, and negative stimuli tend to provoke stronger and quicker emotional, behavioral, and cognitive responses than neutral or positive events [38, 39]. Moreover, previous research shows that negative sentiment has a similar effect on online users' behavior. Postings containing negative sentiment elicit more feedback from other users compared to those with positive sentiment [25]. Negative sentiment of a messages is also more likely to diffuse in subsequent comments compared to positive sentiment [25]. Drawing on these insights, we argue that the tweets containing negative sentiment are more likely to be retweeted. Negative tweets are more likely to provoke action. Retweeting, as a convenient and quick form of showing reaction to a message [7], is more, likely to be triggered when the original messages contains negative sentiment rather than positive or neutral sentiment. This leads to the following hypothesis:

H2: Polarity moderates the effect of sentiment on retweet count of a Twitter message. The effect will be stronger for tweets with negative sentiment than those with positive and neutral sentiment.

Arousal is a state of mobilization. While low arousal is regarded as relaxation, high arousal is regarded as activity [11]. Higher levels of emotional arousal increase action-related behaviors such as moving to help others [15], enhance long term memory of the events [42], and lead to faster response in negotiations [41]. Online users are no exception. A study on of New York Times articles shows that articles containing high arousal sentiment are more likely to be shared by email than those with low arousal sentiment [11]. Hence, we argue that the messages containing high levels of emotional arousal are more likely to elicit reaction from other users. Retweeting a message is a form of reaction to its content [7]. Thus, we expect high-arousal messages to elicit greater reaction from online users in the form of retweets. Thus, we hypothesize that:

H3: The larger the amount of emotional arousal a Twitter message exhibits, the more often it is retweeted.

Information processing is a prerequisite to information diffusion. People tend to process information before disseminating them. Previous research suggests that lexical decision response times

represent ease of processing of the information [57]. In other words, words with lower response times are processed more easily. Similarly, emotional facilitation theory proposes that words with greater meaning are processed faster than those with less meaning [58]. Drawing on these insights, we argue that people are more likely to spread information that is processed more easily and is more meaningful. Hence, we expect the predictors of lexical decision response times to be also important in the context of information diffusion in social media.

Previous research shows that high-arousal negative content is processed faster than neutral content while low-arousal negative content is processed slower than neutral content [17, 18]. We expect content that is processed more quickly to also be diffused more rapidly. Thus, we expect the expression of high-arousal negative emotions in a message to enhance its diffusion in social media while expression of low-arousal negative emotions is expected to deteriorate the diffusion of a message. Consequently, we propose the following hypothesis:

H4: Emotional arousal moderates the effect of negative sentiment on retweet count of a Twitter message. The effect will be positive for tweets with high-arousal sentiment and negative for tweets with low-arousal-sentiment.

## 4. Methodology

### 4.1. Data collection software

C# language was used by the authors to develop software that is capable of collecting tweets. The software uses LinqToTwitter open source framework to connect to Twitter API version 1.1. The developed software allows authors to find tweets related to specific topic, receive a stream of tweets on a specific topic, or download tweets from a specific Twitter account.

We collected 3219 tweets from CNN's Twitter page over a four-month period from October 2013 to January 2014. Table 1 shows the descriptive statistics of the sample. Fifteen percent of the tweets had positive polarity, 40% were negative, and 45% had neutral polarity.

### 4.2. Measurement

The Affective Norms for English Words (ANEW) lexicon was used for analysis of emotional arousal. To calculate emotional arousal of a text using ANEW, we calculated arousal of each word separately and then added the values up for each tweet. A piece of software

was developed by the authors that calculates emotional arousal of a tweet based on the ANEW weights.

**Table 1. Descriptive statistics**

	Minimum	Maximum	Mean	Std. deviation
Arousal	0	43.49	10.32	7.76
Negative Sentiment	-5	-1	-1.67	0.90
Positive Sentiment	1	4	1.28	0.60
Polarity	3	-4	-.40	1.12
Retweet Count	1	147128	293	2668
Total Sentiment	0	5	0.95	1.04

SentiStrength software was be used for measuring total sentiment and emotional valence [59]. The software is free for academic research and has been tested and validated by previous studies [13, 59-63]. SentiStrength is capable of processing different types of information contained in the text including correction of spelling due to repeated letters, analysis of emoticons and booster words, and use of negative words (e.g., not) to flip emotions.

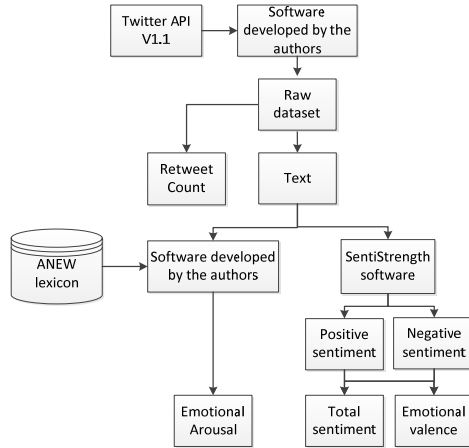
SentiStrength reports two distinct numbers for positive and negative sentiment. The positive number varies from 1 (not positive) to 5 (extremely positive). The negative number varies from -1 (not negative) to -5 (extremely negative). Because both numbers indicate the sentiment of a statement, we use the approach used by Stieglitz and Dang-Xuan [13] to combine the two numbers. We compute the sentiment polarity of each statement by the following formula, which determines the direction of the sentiment (i.e., emotional valence) as well as its strength:

$$polarity = positive\ sentiment + negative\ sentiment,$$

Because positive sentiments vary from 1 to 5 and negative varies from -1 to -5, polarity will have a range of -4 to 4. The other approach to combine the positive and negative numbers is to calculate the total amount of sentiment in a statement regardless of it is positive or negative. To achieve this, the absolute value of positive and negative sentiments should be added up using the following formula:

$$total\ sentiment = (positive\ sentiment - negative\ sentiment) - 2$$

Positive sentiment varies from 1 to 5 and negative sentiment varies from -1 to -5. Thus, total sentiment will have a range of 2 to 10. Hence, we subtracted 2 from (positive – negative) to change the range from [2, 10] to [0, 8]. Figure 2 shows system design and sentiment extraction process.



**Figure 2. System design for sentiment extraction from twitter**

### 4.3. Data Analysis

We first checked the correlation of our items. Arousal, total sentiment, and polarity are significantly correlated. Because we observed relatively high correlations among some variables, we checked the VIF of independent variables. The result of the analysis showed that multicollinearity is not an issue in this study. Table 2 shows the calculated correlations.

We used regression analysis to test the proposed research model. The dependent variable in our model, retweet count, represents nonnegative and integer data and its standard deviation is larger than its mean. Hence, the analysis needs to be adjusted for overdispersion using log-transformation [64]. To analyze hypotheses 1 through 3, we use the following regression model (equation 1):

$$\log(\text{retweet\_count}) = \beta_0 + \beta_1 \text{Total sentiment} + \beta_2 \text{Arousal} + \beta_3 \text{NEGATIVE\_POLARITY} + \beta_4 \text{Total sentiment} * \text{NEGATIVE\_POLARITY} \quad (\text{equation 1})$$

We used negative binomial regression to analyze the model because the dependent variable is a count measure [65]. Negative binomial regression uses log-transformation which addresses the overdispersion problem of the dependent variable. In the above equation, *Arousal* refers to the level of arousal expressed in the message. *NEGATIVE\_POLARITY* is a dummy variable set to 1 if the message has negative polarity and 0 otherwise. We used the interaction term *Total sentiment \* NEGATIVE\_POLARITY* to test the second hypothesis.

To test the fourth hypothesis, we used equation 2 which we analyzed using negative binomial regression. We used polar extremes approach to break

arousal into three levels [66]. *AROUSAL\_LEVEL* is a categorical variable set to 2 if the value for arousal is in the top one-third, to 1 if arousal level is in the bottom one-third, and 0 otherwise. *AROUSAL\_LEVEL* is a factor in this model and the effect of its different levels (i.e., 0, 1, or 2) will be analyzed.

$$\log(\text{retweet\_count}) = \beta_0 + \beta_1 \text{Total sentiment} + \beta_2 \text{AROUSAL\_LEVEL} + \beta_3 \text{NEGATIVE\_POLARITY} + \beta_4 \text{Total sentiment} * \text{NEGATIVE\_POLARITY} * \text{AROUSAL\_LEVEL} \quad (\text{equation 2})$$

**Table 2. Correlations**

	Arousal	Polarity	Retweet	Sentiment
Arousal	1			
Polarity	-0.085**	1		
Retweet	0.022	-0.20	1	
Sentiment	0.268**	-0.347**	0.029	1

\*\* Significant at 0.01

The overall model was significant at  $p < 0.001$ . The results show that sentiment is significantly related to retweet count ( $b = 0.099$ ,  $p < 0.001$ ) providing support for H1. The coefficient for *total sentiment \* NEGATIVE\_POLARITY* was significant ( $b = 0.216$ ,  $p < 0.001$ ), thus we find support for H2. The analysis also shows that emotional arousal significantly predicts retweet behavior ( $b = 0.016$ ,  $p < 0.001$ ), lending support to H3. Analysis of the second equation provides support for our fourth hypothesis. Descriptive statistics show that 15% of the tweets were categorized as high arousal ( $> \mu + \sigma$ ) and 16% were categorized as low arousal ( $< \mu - \sigma$ ). While high-arousal-negative content is more likely to be retweeted ( $b = 0.351$ ,  $p < 0.001$ ), low-arousal-negative content negatively influence retweet count ( $b = -0.157$ ,  $p < 0.02$ ). We also checked the effect of different levels of arousal on diffusion of positive tweets. While positive tweets are more likely to be retweeted than neutral ones ( $b = 0.454$ ,  $p < 0.001$ ), different levels of arousal do not significantly influence the retweet likelihood of positive tweets. The coefficient was insignificant for both high and low arousal levels. Table 3 shows a summary of hypothesis testing.

## 5. Discussion

This study shows how sentiment of a message influences its diffusion in social media. We test two dimensions of sentiment, valence and arousal, and show how they influence diffusion of information in

Twitter. We find that total amount sentiment of a message significantly predicts its diffusion in social media. Drawing upon negativity bias and in line with previous research, we show that negative content is more likely to be retweeted than positive content [13]. We also show that emotional arousal is a significant predictor of diffusion of Twitter messages and high-arousal content is more likely to be retweeted than low arousal content. Our results show that the level of arousal is related to diffusion of negative messages and high-arousal, negative sentiment positively influences diffusion of information on Twitter. However, low-arousal negative sentiment negatively influences retweet count of a message. We also did not find any significant evidence that the level of arousal influences diffusion of positive tweets. Positive tweets are more likely to be retweeted than neutral tweets, but the level of arousal does not seem to affect their diffusion. It seems that emotional arousal becomes effective when a tweet is negative, and, in the case of negative tweets, it would act as expected: high arousal is more likely to lead to action (i.e., retweet) and low arousal is less likely to get users to react. These findings are consistent with those of Lexical Decision Tasks where arousal is an important predictor of response times for negative words. This also indicates that the results from Lexical Decision Task experiments may be generalizable to social media research.

This study contributes to both theory and practice in several ways. From a theoretical perspective, we show valence is not the only important dimension of sentiment affecting information diffusion in social media and other dimensions of sentiment can also significantly predict user behavior. Consequently, we show the significant effect of emotional arousal on information diffusion in social media. This study shows that different dimensions of sentiment may interact with each other and jointly influence user behavior on social media. Finally, we use findings from neuro-science to explain the differences in the effect of emotional arousal on diffusion of positive and negative messages. We argue that information that is processed easily and is more meaningful is more likely to be disseminated in social media. Thus, we show how findings from lexical decision tasks can be used to predict users' retweet behavior. This approach may be used by future research to further explain user behavior on social media.

From a practical perspective, we show how different forms of sentiment influence user behavior on social media. Many organizations try to reach a broader audience on social media and are trying to improve the diffusion of their messages. Many organizations post links to their websites on social

media with the goal of getting the users visit their website. Many of these organizations post a status message with the posted link that is intended to persuade the users to click on the link or forward it to other users. Our findings suggest that selection of the words is very important for organizations. Those responsible for managing the social media accounts of organizations should carefully select the status message that comes with the links they post on social media. Emotional messages are more likely to be dispersed in social media. Moreover, a combination of negative and high arousal sentiment may improve the diffusion of the message.

**Table 3. Hypothesis testing results**

Hypothesis	Hypothesized Relationship	Estimates (Wald Chi-square)	Results
H1	Sentiment → Retweet Count	0.10 (12.75) ***	S
H2	Total sentiment * NEGATIVE_POLARITY → Retweet Count	0.22 (20.99) ***	S
H3	Arousal → Retweet Count	0.02 (37.35) ***	S
H4	Total sentiment * NEGATIVE_POLARITY * [AROUSAL_LEVEL = 2] → Retweet Count	0.351 (49.68) ***	S
	Total sentiment * NEGATIVE_POLARITY * [AROUSAL_LEVEL = 1] → Retweet Count	-0.157 (6.18) *	
*, **, ***: significant at 0.05, 0.01, and 0.001, respectively; S: Supported, NS: Not Supported			

Like any other studies, this study has limitations. First, the software we used in this study is capable of processing different types of text. However, it lacks the processing capability for alternate styles of writing such as sarcasm. While research in the area of natural language processing continues, future studies may provide better insights into determinants of information diffusion in social media by using advanced text analytics techniques. Moreover, our sample lacks language and cultural diversity. The tweets used in this study were collected from CNN Twitter page and are all in English. However, previous research shows that different cultures have differences in terms of how they express emotions. For example, it is known that people in individualistic societies express negative emotions more freely than those living in collectivist ones [67]. Similarly, our results,

collected from a Twitter account of a mainstream broadcaster in the U.S., which has a relatively individualistic culture [68], shows that messages with negative sentiment are more likely to be retweeted. However, the results may not be generalizable to collectivist cultures. On the other hand, language as the medium of communication may play in significant role in how the message is processed by the reader. While our findings regarding the effect of total sentiment and negative emotions on retweet count is similar to those of Stieglitz and Dang-Xuan [13] who examined tweets written in German, the case of emotional arousal may be different. Future research may utilize messages from other cultures and languages, to control for the effect of language and culture on diffusion of information in social media. Future research may also look at the different level of emotions beyond the positive and negative emotion and the differences in expression of emotions between the real life and the virtual space.

## 6. Conclusions

This study investigates the effect of emotions on information diffusion in social media. We find that the level of sentiment a Twitter message carries significantly influences its retweet performance. We also find that negative messages have higher retweet performance than positive ones. Finally, we find that high-arousal, negative sentiment significantly improves information diffusion in social media while low-arousal negative sentiment is negatively related to the diffusion of tweets. This study contributes to the practice by providing individuals, broadcasting agencies, political campaigns, and non-for-profit organizations with strategies to increase their area of impact on social media.

## 7. References

- [1] StatisticsBrain.com *Facebook Statistics*. 2014.
- [2] StatisticsBrain.com *Twitter Statistics*. 2014.
- [3] Naughton, J. *US elections 2012: is Facebook 'the real presidential swing state'?* 2012.
- [4] Kabir, A. *Twiplomacy and the Iran nuclear deal*. 2013.
- [5] Nagarajan, M., H. Purohit, and A.P. Sheth. *A qualitative examination of topical tweet and retweet practices*. in *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*. 2010. Palo Alto, CA.
- [6] Suh, B., et al. *Want to be retweeted? large scale analytics on factors impacting retweet in twitter network*. in *IEEE second international conference on Social computing (socialcom)*. 2010. Los Alamitos, CA: IEEE.
- [7] Recuero, R., R. Araujo, and G. Zago. *How does social capital affect retweets?* in *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*. 2011. Palo Alto, CA.
- [8] Macskassy, S.A. and M. Michelson. *Why do people retweet? Anti-homophily wins the day!* in *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*. 2011. Palo Alto, CA.
- [9] Riordan, M.A. and R.J. Kreuz, *Emotion encoding and interpretation in computer-mediated communication: Reasons for use*. *Computers in Human Behavior*, 2010. **26**(6): p. 1667-1673.
- [10] Walther, J.B. and K.P. D'Addario, *The impacts of emoticons on message interpretation in computer-mediated communication*. *Social Science Computer Review*, 2001. **19**(3): p. 324-347.
- [11] Berger, J. and K.L. Milkman, *What makes online content viral?* *Journal of Marketing Research*, 2012. **49**(2): p. 192-205.
- [12] Salehan, M. and D. Kim, *Predicting the Performance of Online Consumer Reviews: A Sentiment Mining Approach*, in *ICIS 2014 Proceedings*. 2014.
- [13] Stieglitz, S. and L. Dang-Xuan, *Emotions and Information Diffusion in Social Media—Sentiment of Microblogs and Sharing Behavior*. *Journal of Management Information Systems*, 2013. **29**(4): p. 217-248.
- [14] Wundt, W.M., *Grundzüge der physiologischen Psychologie*. Vol. 1. 1908: W. Engelmann.
- [15] Gaertner, S.L. and J.F. Dovidio, *The subtlety of White racism, arousal, and helping behavior*. *Journal of Personality and Social Psychology*, 1977. **35**(10): p. 691.
- [16] Kaufman, B.E., *Emotional arousal as a source of bounded rationality*. *Journal of Economic Behavior & Organization*, 1999. **38**(2): p. 135-144.
- [17] Hofmann, M.J., et al., *Affective processing within 1/10th of a second: High arousal is necessary for early facilitative processing of negative but not positive words*. *Cognitive, Affective, & Behavioral Neuroscience*, 2009. **9**(4): p. 389-397.
- [18] Kuchinke, L., *Implicit and explicit recognition of emotionally valenced words*, in *Department*



- of Educational Science and Psychology. 2007, Freien Universität Berlin.
- [19] Bastos, M.T., R.L.G. Raimundo, and R. Travitzki, *Gatekeeping Twitter: message diffusion in political hashtags*. Media, Culture & Society, 2013. **35**(2): p. 260-270.
- [20] Chevalier, J.A. and D. Mayzlin, *The effect of word of mouth on sales: Online book reviews*. Journal of Marketing Research, 2006. **43**(3): p. 345-354.
- [21] Bakshy, E., et al. *The role of social networks in information diffusion*. in *Proceedings of the 21st international conference on World Wide Web*. 2012. ACM.
- [22] Cha, M., A. Mislove, and K.P. Gummadi. *A measurement-driven analysis of information propagation in the flickr social network*. in *Proceedings of the 18th international conference on World wide web*. 2009. ACM.
- [23] Gruhl, D., et al. *Information diffusion through blogspace*. in *Proceedings of the 13th international conference on World Wide Web*. 2004. ACM.
- [24] Garg, R., M.D. Smith, and R. Telang, *Measuring information diffusion in an online community*. Journal of Management Information Systems, 2011. **28**(2): p. 11-38.
- [25] Stieglitz, S. and L. Dang-Xuan. *Impact and Diffusion of Sentiment in Public Communication on Facebook*. in *ECIS 2012 Proceedings*. 2012.
- [26] Yang, J. and S. Counts, *Predicting the Speed, Scale, and Range of Information Diffusion in Twitter*. ICWSM, 2010. **10**: p. 355-358.
- [27] Romero, D.M., B. Meeder, and J. Kleinberg. *Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter*. in *Proceedings of the 20th international conference on World wide web*. 2011. ACM.
- [28] Lee, C., et al. *Finding influentials based on the temporal order of information adoption in twitter*. in *Proceedings of the 19th international conference on World wide web*. 2010. ACM.
- [29] Bakshy, E., et al. *Everyone's an influencer: quantifying influence on twitter*. in *Proceedings of the fourth ACM international conference on Web search and data mining*. 2011. ACM.
- [30] Lerman, K. and R. Ghosh, *Information Contagion: An Empirical Study of the Spread of News on Digg and Twitter Social Networks*. ICWSM, 2010. **10**: p. 90-97.
- [31] Yang, J. and J. Leskovec. *Modeling information diffusion in implicit networks*. in *Data Mining (ICDM), 2010 IEEE 10th International Conference on*. 2010. IEEE.
- [32] Harris, R.B. and D. Paradice, *An investigation of the computer-mediated communication of emotions*. Journal of Applied Sciences Research, 2007. **3**(12): p. 2081-2090.
- [33] Sen, S. and D. Lerman, *Why are you telling me this? An examination into negative consumer reviews on the web*. Journal of Interactive Marketing, 2007. **21**(4): p. 76-94.
- [34] Schindler, R.M. and B. Bickart, *Perceived helpfulness of online consumer reviews: The role of message content and style*. Journal of Consumer Behaviour, 2012. **11**(3): p. 234-243.
- [35] Bai, X., *Predicting consumer sentiments from online text*. Decision Support Systems, 2011. **50**(4): p. 732-742.
- [36] Mehrabian, A. and J.A. Russell, *An approach to environmental psychology*. 1974: the MIT Press.
- [37] Plutchik, R., *Emotion: A psychoevolutionary synthesis*. 1980: Harper & Row New York.
- [38] Rozin, P. and E.B. Royzman, *Negativity bias, negativity dominance, and contagion*. Personality and social psychology review, 2001. **5**(4): p. 296-320.
- [39] Baumeister, R.F., et al., *Bad is stronger than good*. Review of general psychology, 2001. **5**(4): p. 323.
- [40] Heilman, K., *The neurobiology of emotional experience*, in *The neuropsychiatry of limbic and subcortical disorders*. 1997. p. 133-142.
- [41] Brooks, A.W. and M.E. Schweitzer, *Can Nervous Nelly negotiate? How anxiety causes negotiators to make low first offers, exit early, and earn less profit*. Organizational Behavior and Human Decision Processes, 2011. **115**(1): p. 43-54.
- [42] Cahill, L. and J.L. McGaugh, *A novel demonstration of enhanced memory associated with emotional arousal*. Consciousness and cognition, 1995. **4**(4): p. 410-421.
- [43] Briesemeister, B.B., L. Kuchinke, and A.M. Jacobs, *Discrete emotion effects on lexical decision response times*. PloS one, 2011. **6**(8): p. e23743.
- [44] Bradley, M.M. and P.J. Lang, *Affective norms for English words (ANEW): Instruction manual and affective ratings*. 1999, The Center for Research in Psychophysiology, University of Florida.

- [45] Bellezza, F.S., A.G. Greenwald, and M.R. Banaji, *Words high and low in pleasantness as rated by male and female college students*. Behavior Research Methods, Instruments, & Computers, 1986. **18**(3): p. 299-303.
- [46] Bradley, M.M. and P.J. Lang, *Affective norms for English words (ANEW): Instruction manual and affective ratings*. 2010, Technical Report C-1, The Center for Research in Psychophysiology, University of Florida.
- [47] Yu, H. and V. Hatzivassiloglou. *Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences*. in *Proceedings of the 2003 conference on Empirical methods in natural language processing*. 2003. Association for Computational Linguistics.
- [48] Bradley, M.M., et al., *Emotion and motivation I: defensive and appetitive reactions in picture processing*. Emotion, 2001. **1**(3): p. 276.
- [49] Jegadeesh, N. and D. Wu, *Word power: A new approach for content analysis*. Journal of Financial Economics, 2013. **110**(3): p. 712-729.
- [50] Weymar, M., et al., *When fear forms memories: Threat of shock and brain potentials during encoding and recognition*. Cortex, 2013. **49**(3): p. 819-826.
- [51] Sherdell, L., C.E. Waugh, and I.H. Gotlib, *Anticipatory pleasure predicts motivation for reward in major depression*. Journal of abnormal psychology, 2012. **121**(1): p. 51.
- [52] Naveed, N., et al. *Bad news travel fast: A content-based analysis of interestingness on twitter*. in *Proceedings of the 3rd International Web Science Conference*. 2011. ACM.
- [53] Whissell, C., *Using the revised dictionary of affect in language to quantify the emotional undertones of samples of natural language 1, 2*. Psychological reports, 2009. **105**(2): p. 509-521.
- [54] Whissell, C., *Whissell's dictionary of affect in language: Technical manual and user's guide*. 2010, Laurentian University.
- [55] Huffaker, D., *Dimensions of leadership and social influence in online communities*. Human Communication Research, 2010. **36**(4): p. 593-617.
- [56] Joyce, E. and R.E. Kraut, *Predicting continued participation in newsgroups*. Journal of Computer-Mediated Communication, 2006. **11**(3): p. 723-747.
- [57] Ferraro, F.R., K. Christopherson, and J. Douglas, *Lexical decision task performance in blood-fearful and spider-fearful individuals*. Current Psychology, 2006. **25**(2): p. 132-143.
- [58] Balota, D.A. and J.I. Chumbley, *Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage*. Journal of Experimental Psychology: Human perception and performance, 1984. **10**(3): p. 340.
- [59] Thelwall, M., et al., *Sentiment strength detection in short informal text*. Journal of the American Society for Information Science and Technology, 2010. **61**(12): p. 2544-2558.
- [60] Thelwall, M., K. Buckley, and G. Paltoglou, *Sentiment strength detection for the social web*. Journal of the American Society for Information Science and Technology, 2012. **63**(1): p. 163-173.
- [61] Thelwall, M. and K. Buckley, *Topic-based sentiment analysis for the social web: The role of mood and issue-related words*. Journal of the American Society for Information Science and Technology, 2013. **64**(8).
- [62] Garcia, D. and F. Schweitzer. *Emotions in Product Reviews-Empirics and Models*. in *2011 IEEE International Conference on Privacy, Security, Risk, and Trust, and IEEE International Conference on Social Computing*. 2011. IEEE.
- [63] Gruz, A., S. Doiron, and P. Mai. *Is happiness contagious online? A case of Twitter and the 2010 Winter Olympics*. in *Proceedings of the 44th Hawaii International Conference on System Sciences*. 2011. IEEE.
- [64] Cameron, A.C. and P.K. Trivedi, *Regression Analysis of Count Data*. 2013: Cambridge University Press.
- [65] Hilbe, J.M., *Negative Binomial Regression*. 2011: Cambridge University Press.
- [66] George, B. and V. Prybutok, *Development of a polar extreme method for use in partial least squares SEM*. Quality & Quantity, 2014. **49**(2): p. 1-18.
- [67] Takahashi, K., et al., *Commonalities and differences in close relationships among the Americans and Japanese: A comparison by the individualism/collectivism concept*. International Journal of Behavioral Development, 2002. **26**(5): p. 453-465.
- [68] Hofstede, G., *Culture's consequences: International differences in work-related values*. Vol. 5. 1984: Sage.